

LIST DECODING OF NOISY REED-MULLER-LIKE CODES

A. ROBERT CALDERBANK, ANNA C. GILBERT, AND MARTIN J. STRAUSS

ABSTRACT. Coding theory has played a central role in the development of computer science. One critical point of interaction is decoding error-correcting codes. First- and second-order Reed-Muller (RM(1) and RM(2), respectively) codes are two fundamental error-correcting codes which arise in communication as well as in probabilistically-checkable proofs and learning. In this paper, we take the first steps toward extending the quick randomized decoding tools of RM(1) into the realm of quadratic binary and, equivalently, \mathbb{Z}_4 codes. Our main algorithmic result is an extension of the RM(1) techniques from Goldreich-Levin and Kushilevitz-Mansour algorithms [GL89, KM91] to the *Hankel* code [CGL⁺05], a code between RM(1) and RM(2). That is, given signal s of length N , we find a list that is a superset of all Hankel codewords φ with $|\langle s, \varphi \rangle|^2 \geq (1/k) \|s\|^2$, in time $\text{poly}(k, \log(N))$. We then turn our attention to the widely-studied Kerdock codes. We give a new and simple formulation of a known Kerdock code as a subcode of the Hankel code. We then get two immediate corollaries. First, our new Hankel list-decoding algorithm covers subcodes, including the new Kerdock construction, so we can list-decode Kerdock, too. Furthermore, exploiting the fact that dot products of distinct Kerdock vectors have small magnitude, we get a quick algorithm for finding a sparse Kerdock approximation. That is, for k small compared with $1/\sqrt{N}$ and for $\epsilon > 0$, we find, in time $\text{poly}(k \log(N)/\epsilon)$, a k -Kerdock-term approximation \tilde{s} to s with Euclidean error at most the factor $(1 + \epsilon + O(k^2/\sqrt{N}))$ times that of the best such approximation.

1. INTRODUCTION

Coding theory and computation have enjoyed a long and fruitful interaction. Decoding a received codeword is inherently an algorithmic problem and, conversely, codes have been used as key components of algorithms for many purposes, including pseudorandomness, probabilistically checkable proofs, learning, and cryptography. The computational view of codes can also provide important insights for coding theory and code construction. See [Sud, Sud01] and the references therein for a sample of this fruitful interaction.

Because decoding is inherently an algorithmic problem, it is natural to analyze the computational cost of decoding a received codeword. We can quantify how much time and space we need to decode a vector which has been corrupted according to a variety of noise models. In this paper, we are interested in how many samples of the received codeword are necessary for decoding, how much noise we can tolerate in the input, and how quickly we can decode using just a few random samples in the presence of this noise.

The first- and second-order binary Reed-Muller codes RM(1) and RM(2) are fundamental in the study of codes and their applications to algorithms. A RM(1) codeword of dimension n can be regarded as a binary linear function on n variables and a RM(2) codeword is a quadratic function on n variables. As such, they are fundamental expressive classes, used in proofs and learning as well as error-free communication.

A. Robert Calderbank is with the Department of Mathematics, Princeton University, Princeton, NJ 08544. Anna C. Gilbert is with the Department of Mathematics, The University of Michigan at Ann Arbor, 2074 East Hall, 530 Church St., Ann Arbor, MI 48109-1043. Martin J. Strauss is jointly appointed with the Department of Mathematics and the Department of Electrical Engineering and Computer Science at The University of Michigan. Calderbank has been supported by DARPA-ONR N00173-06-1-G006. Gilbert and Strauss have been supported by NSF DMS 0354600 and DARPA-ONR N66001-06-1-2011.

Binary RM(1), in particular, admits highly efficient algorithms for decoding, even in the presence of noise. We are interested in a form of decoding that has appeared many times before with various names, and that we call *Euclidean List Decoding*. The first quick algorithms for Euclidean list decoding of RM(1) are in [GL89, KM91]. Given a (multiplicatively-written) linear function $f : \mathbb{Z}_2^n \rightarrow (\pm 1, \cdot)$, one can recover f by querying its value on just $\text{poly}(n)$ values of its graph, instead of all 2^n values. Furthermore, the decoding succeeds even in the presence of a lot of noise; *i.e.*, if the noise ν is orthogonal to the signal f and if we assume only that $\|f\|^2 \geq (1/k) \|\nu\|^2$, then the algorithm on $f + \nu$ takes time polynomial in kn and returns a (short) list of possible f 's. See [Sud00] for a discussion of list decoding algorithms and their applications. We note that this problem can be solved more generally using nearest-neighbor data structures [Ind00], but the general solution requires space and preprocessing time $N = 2^n$, which we want to avoid.

While the available techniques for RM(1) make it useful in many applications, RM(1) is limited in several important ways compared with RM(2). First, there are only 2^n RM(1) codewords, while there are approximately $2^{n^2/2}$ RM(2) codewords, so, quantitatively, RM(2) is more expressive. But there are important structural differences, as well. When used to express a concept or to code a computation, an RM(1) codeword as a function considers its variables one at a time, while RM(2) codewords consider their variables in pairs. First-order Reed-Muller codewords form an orthonormal basis, while RM(2) forms a highly redundant *dictionary*—a collection of more than N vectors spanning a vector space of dimension N —that is potentially much more useful for lossy compression. When used as a pseudorandom number generator, RM(1) provides a family of 3-wise independent random variables and RM(2) provides a family of 7-wise random variables.¹ Because of the extra expressiveness of RM(2), however, many tools from the first order theory do not apply. For example, we do not know how to recover a RM(2) vector in the presence of noise unless the noise is slight [AKK⁺03].

In this paper, we take the first steps toward extending the decoding tools of RM(1) into the realm of quadratic binary (and, equivalently, \mathbb{Z}_4) codes. We show how to recover *Hankel* codewords [CGL⁺05] efficiently in the presence of noise, giving a result analogous to what one can do with RM(1) up to a polynomial in the parameters. The Hankel code is the union of *cosets* of RM(1), *i.e.*, $\bigcup_{\varphi \in Q} \varphi \text{RM}(1)$ for some Q of size q , so that Hankel can be regarded as the union of q orthonormal bases, each equivalent to RM(1). It follows immediately that one can use the [KM91] algorithm q times to do list-decoding over the union of q equivalent copies of RM(1), but only at time cost q times the cost of one instance of the algorithm in [KM91]. Hankel consists of $q = \Theta(N^2)$ copies of RM(1), however, so the cost of such a trivial algorithm would be prohibitive. By contrast, we list-decode Hankel in total time $\text{poly}(k, \log(N))$. Such efficient list-decoding is possible only by confluence of the choice of dictionary (Hankel) and the algorithm, and represents an important way in which our contribution is significant.

We also give a new, simple construction of a code in the well-studied class of *Kerdock* codes. Our Kerdock construction \mathcal{K} is a subcode of the Hankel code \mathcal{H} , which implies immediately that our Hankel list-decoding algorithm applies also to our Kerdock construction. Thus we have $\text{RM}(1) \subseteq \mathcal{K} \subseteq \mathcal{H} \subseteq \text{RM}(2)$. While Kerdock and Hankel are still in some important respects more limited than RM(2), they are great improvements over RM(1). For example, a random codeword from a Kerdock code (and, therefore, from the Hankel code) provides a family of 5-wise independent random variables. Each Kerdock code has N^2 vectors and the Hankel code has $\Theta(N^3)$ vectors, compared with $\Theta(N)$ for RM(1) and $2^{\Theta(\log^2(N))}$ for RM(2). Kerdock represents a substantial,

¹That is, if we fix any three indices y_1, y_2, y_3 into an unknown codeword φ and then choose an RM(1) codeword φ at random, the random variables $\varphi(y_1), \varphi(y_2), \varphi(y_3)$ are jointly independent. If we choose an RM(2) codeword at random, any 7 positions are independent.

well-studied family of quadratic functions with advantages over RM(1) in the areas of coding theory [HKC⁺94], radar signaling [HCM06], and spread-spectrum communication.

Finally, the previous work in [TGMS03, GMS03] demonstrates that we can use a fast list-decoding algorithm for to find a *sparse representation* efficiently. That is, exploiting the fact that dot products of distinct Kerdock vectors have small magnitude, we get a quick algorithm for finding a sparse Kerdock approximation. More specifically, for any $k < 1/(6\sqrt{N})$ and any $\epsilon > 0$, we can find, in time $\text{poly}(k, \log(N), 1/\epsilon)$, a k -Kerdock-term approximation \tilde{s} to s with Euclidean error at most the factor $(1 + \epsilon + O(k^2/\sqrt{N}))$ times that of the best such approximation.

This paper is organized as follows. In Section 2, we give preliminaries about finite fields, Reed-Muller codes, and Kerdock codes. We also include a discussion of related work. In Section 3, we give a new, computational construction of a Kerdock code, as a subcode of the Hankel code. In Section 4, we give our algorithm for fast list decoding of the Hankel code. In Section 5, we give corollaries of our main result concerning list-decoding and sparse recovery of Kerdock codes, as well as indications about directions for improvement.

2. PRELIMINARIES

2.1. Finite fields. To outline the setting in which Kerdock codes are defined, we begin with the definition of finite fields and the algebra we perform over these fields. Let $h(t)$ be a polynomial of degree n over \mathbb{Z}_2 that is primitive, *i.e.*, $h(t)$ does not divide $t^k - 1$ for any $k < 2^n - 1$. Because h is a primitive (and hence, irreducible) polynomial, it has no non-trivial factorization.

The ring of polynomials $\mathbb{Z}_2[t]$ modulo h , $\mathbb{Z}_2[t]/h$, forms a field of 2^n elements. We denote this field $\mathbb{F}(2^n)$. The polynomial $\xi(t) = t$ is a (multiplicative) generator of the field; thus, the set $\{1, \xi, \xi^2, \dots, \xi^{2^n-1}\}$ enumerates the non-zero elements of the field. Additively, the field $\mathbb{F}(2^n)$ is a vector space \mathbb{Z}_2^n over \mathbb{Z}_2 of dimension n with basis $\{1, \xi, \xi^2, \dots, \xi^{n-1}\}$. It is also a quotient vector space of \mathbb{Z}^n . When we want to emphasize the vector formulation of a field element α , we write $[\alpha]$ for a column vector. Thus $[1], [\xi], [\xi^2], \dots, [\xi^{n-1}]$ are the canonical basis vectors. Below, we will often want to consider these $\{0, 1\}$ -valued vectors to be in $\mathbb{Z}_2^n, \mathbb{Z}_4^n$, or \mathbb{Z}^n for the purposes of dot products. We will write, *e.g.*, $i[y]^T Q[y] + 2\ell^T[y]$, where y is a field element, Q is a $\{0, 1\}$ -valued matrix, and ℓ is a $\{0, 1\}$ -valued vector. Note that all the arithmetic in the exponent can be done over \mathbb{Z} , where $[y]$ is a $\{0, 1\}$ -valued vector. Since the exponent is an exponent of i , arithmetic can equivalently be done mod 4. Finally, since 2 multiplies $\ell^T[y]$, the dot product of ℓ and $[y]$ can be performed mod 2. For any $x \in \mathbb{F}(2^n)$, we have $x^{2^n} = x$ so that $\sqrt{x} = x^{2^{n-1}}$. Because 2 is congruent to 0 mod 2, we have $(x + y)^2 = x^2 + y^2$ for any $x, y \in \mathbb{F}(2^n)$ and, by repeated squaring, $(x + y)^{2^j} = x^{2^j} + y^{2^j}$.

The trace of an element $x \in \mathbb{F}(2^n)$ is an important quantity we use in defining and constructing Kerdock codes.

Definition 1. *The trace of $x \in \mathbb{F}(2^n)$, $\text{Tr}(x)$, is defined to be*

$$\text{Tr}(x) = \sum_{0 \leq j < n} x^{2^j} = x + x^2 + \dots + x^{2^{n-1}}.$$

The following lemma gives the properties we need of the trace map. We give the simple proof for completeness.

Lemma 2. *We have*

- For $x, y \in \mathbb{F}(2^n)$ and $a, b \in \mathbb{F}(2)$, we have $\text{Tr}(ax + by) = a\text{Tr}(x) + b\text{Tr}(y)$.
- The image of Tr is in \mathbb{Z}_2 .
- The trace is not identically 0.

Proof. (Repeated) squaring of an element is a linear operator, so $\text{Tr}(ax + by) = a\text{Tr}(x) + b\text{Tr}(y)$.

Again by linearity of squaring, $\text{Tr}(x^2) = \text{Tr}(x)^2$. Since $x^{2^n} = x$, we have $\text{Tr}(x) = \text{Tr}(x^2) = \text{Tr}(x)^2$. Thus $\text{Tr}(x)$ satisfies $y = y^2$, whence $\text{Tr}(x) \in \mathbb{F}(2)$. For n odd, $\text{Tr}(1) = 1$. (Lemma 11 shows that $\text{Tr} \neq 0$ for even n , as well.) ■

Thus Tr is an additive homomorphism from the big field $\mathbb{F}(2^n)$ to the *prime subfield* $\mathbb{F}(2) = \mathbb{Z}_2$, so $\text{Tr}(x) = 0$ for exactly half of the field elements. It is not necessarily true that $\text{Tr}(xy) = \text{Tr}(x)\text{Tr}(y)$. Finally, note that $\text{Tr}(1)$ is 0 or 1 if n is even or odd, respectively.

2.2. Definitions of $\text{RM}(1,n)$ and $\text{RM}(2,n)$. We review the definitions of the two codes, first- and second-order Reed-Muller codes ($\text{RM}(1,n)$ and $\text{RM}(2,n)$, respectively), which sandwich Kerdock codes. Fix a parameter n .

Definition 3. Let $\ell \in \mathbb{Z}_2^n$ be a binary vector of length n and let $\epsilon \in \mathbb{Z}_2$. The first-order Reed-Muller code $\text{RM}(1,n)$ of length $N = 2^n$ is defined as a set of vectors $v_{\ell,\epsilon}$ indexed by ℓ and ϵ . For each code word $v_{\ell,\epsilon}$ at position $[y] \in \mathbb{Z}_2^n$ is given by

$$v_{\ell,\epsilon}([y]) = 2(\ell^T[y] + \epsilon) \pmod{4}.$$

The exponentiated form of $\text{RM}(1,n)$ is given by

$$\varphi_{\ell,\epsilon}([y]) = \frac{1}{\sqrt{N}} i^{2(\ell^T[y] + \epsilon)} = \frac{(-1)^{\ell^T[y] + \epsilon}}{\sqrt{N}}.$$

We normalize the codevectors by \sqrt{N} in the exponentiated form to obtain unit vectors.

Definition 4. Let Q be an $n \times n$ symmetric matrix over \mathbb{Z}_2 , let $\ell \in \mathbb{Z}_2^n$ be a binary vector of length n , and let $\epsilon \in \mathbb{Z}_4$. The second-order Reed-Muller code $\text{RM}(2,n)$ of length $N = 2^n$ is defined as a set of vectors $w_{Q,\ell,\epsilon}$ indexed by Q , ℓ , and ϵ . Each codeword $w_{Q,\ell,\epsilon}$ at position $[y] \in \mathbb{Z}_2^n$ is given by

$$w_{Q,\ell,\epsilon}([y]) = ([y]^T Q [y] + 2\ell^T [y] + \epsilon) \pmod{4}.$$

The exponentiated form of $\text{RM}(2,n)$ is given by

$$\varphi_{Q,\ell,\epsilon}([y]) = \frac{1}{\sqrt{N}} i^{[y]^T Q [y] + 2\ell^T [y] + \epsilon}.$$

Below, we will sometimes abbreviate the index (Q, ℓ, ϵ) as λ , so that $\varphi_{Q,\ell,\epsilon} = \varphi_\lambda$. Again, we normalize the codevectors in the exponentiated form so they are unit vectors. Observe that if $Q = 0$, then the subset of $\text{RM}(2,n)$ codewords given by $w_{0,\ell,\epsilon}$ are, in fact, $\text{RM}(1,n)$ codewords. We frequently drop the index ϵ since i^ϵ represents a unit factor that can be absorbed into a more general coefficient $c_{Q,\ell}$ of $c_{Q,\ell}\varphi_{Q,\ell}$.

In other literature, both $\text{RM}(1)$ and $\text{RM}(2)$ are presented as binary codes. Our theory can be formulated for both \mathbb{Z}_2 and \mathbb{Z}_4 , but we stick to \mathbb{Z}_4 after giving the equivalence between previous work and ours. We will consider $\text{RM}(1)$ and $\text{RM}(2)$ over \mathbb{Z}_4 , as above, since the Kerdock codes are most natural over \mathbb{Z}_4 —they are nonlinear binary codes but linear over \mathbb{Z}_4 .

We say that a code with entries in \mathbb{Z}_4 is a \mathbb{Z}_4 -code while one with entries in \mathbb{Z}_2 is a \mathbb{Z}_2 -code. The two previous definitions of $\text{RM}(1,n)$ and $\text{RM}(2,n)$ both result in \mathbb{Z}_4 -codes. The \mathbb{Z}_2 Reed-Muller codes may be more familiar to the reader and we often want to relate a \mathbb{Z}_4 -code to a \mathbb{Z}_2 -code. We do so via the Gray map.

Definition 5. The Gray map, $\text{gr} : \mathbb{Z}_4 \rightarrow \mathbb{Z}_2^2$, is given by

$$\text{gr}(0) = 00, \quad \text{gr}(1) = 01, \quad \text{gr}(2) = 11 \quad \text{and} \quad \text{gr}(3) = 10.$$

We sometimes use the exponential version, from $\{\pm 1, \pm i\}$ to $(\pm 1)^2$, given by

$$\text{gr}(+1) = (+1, +1), \quad \text{gr}(+i) = (+1, -1), \quad \text{gr}(-1) = (-1, -1), \quad \text{and} \quad \text{gr}(-i) = (-1, +1).$$

Further overloading notation, $\text{gr} : \mathbb{Z}_4^N \rightarrow \mathbb{Z}_2^{N \times 2}$ is gotten by applying the Gray map to each of N elements in a vector in \mathbb{Z}_4^N , getting N elements in \mathbb{Z}_2^2 , and similarly for the exponential versions.

Equivalently, one can transform Q , a \mathbb{Z}_4 -valued quadratic form on \mathbb{Z}_2^n , to M , a \mathbb{Z}_2 -valued quadratic form on \mathbb{Z}_2^{n+1} . The quadratic form Q is an $n \times n$ binary symmetric matrix while M is an $(n+1) \times (n+1)$ binary skew symmetric matrix—that is, M has zero diagonal. Let the row vector d_Q be the diagonal of Q . Then Calderbank et al. [CCKS97] show that the correspondence between binary symmetric matrices Q and binary skew symmetric matrices M is given by

$$M = \begin{pmatrix} 0 & d_Q^T \\ d_Q & d_Q d_Q^T + Q \end{pmatrix}, \quad (1)$$

where the “extra” bit in the top row and left column is used as an index into the two outputs of the Gray map. This correspondence is not linear but it is rank preserving in the sense that if M has rank $n+1-2j$ then Q has rank $n+1-2j$ or $n-2j$ for any integer j , $0 \leq j < (n-1)/2$.

In summary, the following commutative diagram relates codewords and codeword labels in the \mathbb{Z}_2 and \mathbb{Z}_4 formulations:

$$\begin{array}{ccc} \mathbb{Z}_4 \text{ label} & \xrightarrow{(1)} & \mathbb{Z}_2 \text{ label} \\ \downarrow & & \downarrow \\ \mathbb{Z}_4 \text{ codeword} & \xrightarrow{\text{Gray map}} & \mathbb{Z}_2 \text{ codeword} \end{array}$$

The following theorem of Calderbank et al. [CCKS97] relates the rank of the binary symmetric matrices Q_1 and Q_2 to the magnitude of the dot product between two codewords generated with the respective matrices.

Theorem 6. *Let Q_1 and Q_2 be binary symmetric $n \times n$ matrices and let $\varphi_{Q_1, \ell_1, \epsilon_1}$ and $\varphi_{Q_2, \ell_2, \epsilon_2}$ be distinct exponentiated \mathbb{Z}_4 -RM(2, n) codewords. If $\text{Rank}(Q_1 - Q_2) = R$, then*

$$|\langle \varphi_{Q_1, \ell_1, \epsilon_1}, \varphi_{Q_2, \ell_2, \epsilon_2} \rangle| \in \{0, 2^{-R/2}\}.$$

In particular, if $\ell_1 = \ell_2$, then the magnitude of the dot product is $2^{-R/2}$.

2.3. Definition of Kerdock codes. A Kerdock code is associated with a Kerdock set of matrices. The definition of the latter is non-constructive.

Definition 7. A Kerdock set \mathcal{K} is a set of $n \times n$ binary symmetric matrices, including zero, of size n such that for any distinct $P_1, P_2 \in \mathcal{K}$, the rank of $(P_1 + P_2)$ over $\mathbb{F}(2)$ is n .

In particular, any non-zero $P \in \mathcal{K}$ has full rank. We take these matrices P to be quadratic forms over \mathbb{Z}_4 . Each Kerdock set has size at most $N = 2^n$, since distinct elements of a Kerdock set must have distinct top rows. In fact, Kerdock sets can achieve maximal size (see below).

Definition 8. A Kerdock code $K(n)$ of length $N = 2^n$ is defined as a set of vectors $c_{P, \ell, \epsilon}$, indexed by P , ℓ , and ϵ . Each codeword $c_{P, \ell, \epsilon}$ at position $[y] \in \mathbb{Z}_2^n$ is given by

$$c_{P, \ell, \epsilon}([y]) = ([y]^T P [y] + \ell^T [y] + \epsilon) \mod 4$$

where $P \in \mathcal{K}$ comes from a Kerdock set.

2.4. Related Work. The work most closely related to our decoding algorithm is that of [GL89, KM91], which was already discussed. Similar sparse decoding of the Fourier basis (over \mathbb{Z}_N , not \mathbb{Z}_2^n) was given in [Man95, GGI⁺02, AGS03]. Other work on local testing of codes [KL05] focuses on limiting the number of samples, but not the runtime. Still other work on list-decoding of Reed-Muller codes [Sud01] focuses on large alphabets, whereas we work over \mathbb{Z}_2 . The problem of *testing* low-degree polynomials [AS97] is different from *decoding*, which is what we do for special quadratic polynomials. We note that [AKK⁺03], in addition to giving lower bounds on the number of samples for testing binary Reed-Muller codes, also give a decoding algorithm for a single Reed-Muller vector in the presence of very small noise.

As for construction of Kerdock codes, the history is as follows. Kerdock codes were first defined [MS77] non-constructively in terms of the allowable quadratic forms. Later, in the breakthrough paper [CKKS97], the authors give algebraic constructions of Kerdock codes that provide a rich set of symmetries, but the algebra included theory somewhat beyond finite fields. Independent of and somewhat earlier than the publication of our work, a construction of a Kerdock code similar to ours is given in [HSP06]. Both the construction in [HSP06] and our construction here are isomorphic, in some sense, to a construction in [CKKS97]. We believe our construction is a bit simpler than [HSP06]—indeed, to get the Hankel structure we need here, it is simpler for us to give Definition 9 (below) from scratch than to adapt the construction in [HSP06]. As additional value beyond [HSP06], we also contribute a self-contained proof of correctness of the construction, simplifying the proof in [CKKS97] ([HSP06] gives offers no new proof of correctness). We also give an important new characterization of the construction, Lemma 15.

3. NEW DEFINITION OF KERDOCK CODES

In this section, we present a construction of Kerdock codes.

3.1. Kerdock matrices. In each construct, a Kerdock code of length $N = 2^n$ is a subset of $\text{RM}(2, n)$ code $\{\varphi_{Q, \ell}\}$ satisfying an appropriate restriction on the binary symmetric matrix Q . We call these matrices Kerdock matrices. Roughly speaking, they are a restricted set of binary Hankel² matrices where the top row of the Hankel matrix consists of arbitrary entries and each of the remaining reverse diagonals is gotten from a fixed linear combination of the previous n reverse diagonals.

Let $h(t) = h_0 + h_1 t + \cdots h_{n-1} t^{n-1} + t^n$ be a primitive polynomial over \mathbb{Z}_2 of degree n . The coefficients of this polynomial are the coefficients in our fixed linear mapping.

Definition 9. An $n \times n$ linear-feedback-Kerdock matrix (briefly, lf-Kerdock matrix) is a Hankel matrix where the top row of the matrix a_0, a_1, \dots, a_{n-1} consists of n arbitrary values in \mathbb{Z}_2 and the j th reverse diagonal parameter for $j \geq n$ is a fixed linear combination of the previous n reverse diagonal parameters, given by

$$a_j = \sum_{0 \leq \ell < n} a_{j-n+\ell} h_\ell.$$

(See Section 3.2 for an example.) We denote by \mathcal{K} the set of lf-Kerdock matrices. Next, we provide what turns out to be an equivalent definition of lf-Kerdock matrices, called trace-Kerdock matrices.

Definition 10. An n -by- n trace-Kerdock matrix K_α is the matrix whose (j, k) position is $\text{Tr}(\alpha \xi^{j+k})$ for some α in $\mathbb{F}(2^n)$.

²A Hankel matrix is constant along reverse diagonals.

Note that the set of trace-Kerdock *matrices* is \mathbb{Z}_2 -linear, meaning the sum (mod 2) of two trace-Kerdock matrices is itself a trace-Kerdock matrix, by additivity of the trace. Kerdock *codes*, however, are \mathbb{Z}_4 -linear but *not* \mathbb{Z}_2 -linear.

Lemma 11. *Trace-Kerdock matrices have full rank.*

Proof. Given trace-Kerdock matrix K_α for $\alpha \neq 0$, regard it as a matrix over $\mathbb{F}(2^n)$. Since K_α consists of 0's and 1's, the determinant of K_α over $\mathbb{F}(2^n)$ is the same as the determinant over \mathbb{Z}_2 .

The matrix K_α factors as $K_\alpha = V^T D_\alpha V$, where

$$V = \begin{pmatrix} 1 & \xi & \xi^2 & \xi^3 & \dots & \xi^{n-1} \\ 1 & \xi^2 & \xi^4 & \xi^6 & \dots & \xi^{2(n-1)} \\ 1 & \xi^4 & \xi^8 & \xi^{12} & \dots & \xi^{4(n-1)} \\ \vdots & & & & & \\ 1 & \xi^{2^k} & \xi^{2^k \cdot 2} & \xi^{2^k \cdot 3} & \dots & \xi^{2^k \cdot (n-1)} \\ \vdots & & & & & \\ 1 & \xi^{2^{n-1}} & \xi^{2^{n-1} \cdot 2} & \xi^{2^{n-1} \cdot 3} & \dots & \xi^{2^{n-1} \cdot (n-1)} \end{pmatrix}$$

is vandermonde and $D_\alpha = \text{diag}(\alpha, \alpha^2, \alpha^4, \alpha^8, \dots, \alpha^{2^{n-1}})$. Over the big field,

$$\det(D_\alpha) = \alpha^{1+2+4+\dots+2^{n-1}} = \alpha^{2^n-1} = 1$$

and the vandermonde parameters ξ, ξ^2, ξ^4, \dots are distinct, so V is non-singular. It follows that $\det(K_\alpha) \neq 0$ over the field, so $\det(K_\alpha) = 1$. ■

Note that the factorization $K_\alpha = V^T D_\alpha V$ also shows that $K_x J K_y J = K_{xy} J$, where $J = (V^T V)^{-1} = K_1^{-1}$. Thus the map $x \mapsto K_x J$ is a non-trivial multiplicative homomorphism from field elements to matrices. Since squaring is linear, $x \mapsto D_x$ and, so, $x \mapsto K_x J = V^T D_x (V^T)^{-1}$ are additive homomorphisms. It follows that $x \mapsto K_x J$ is a field homomorphism, so, in conjunction with the matrix J , the trace-Kerdock matrices can be regarded as field elements.

Lemma 12. *Every trace-Kerdock matrix is a lf-Kerdock matrix.*

Proof. Fix a trace-Kerdock matrix K_α . It is Hankel by inspection, since the matrix entry $K_\alpha(j, k) = \text{Tr}(\alpha \xi^{j+k})$ depends only on $j+k$. Note that the first n diagonal parameters are given by

$$\text{Tr}(\alpha), \text{Tr}(\alpha \xi), \text{Tr}(\alpha \xi^2), \dots, \text{Tr}(\alpha \xi^{n-1}).$$

Fix j and k with $j < n, k < n$, and $j+k \geq n$. Then the $j+k$ reverse diagonal of K_α , which can be taken mod 2, is $[\xi^j]^T K_\alpha [\xi^k] = \text{Tr}(\alpha \xi^{j+k})$. Using additivity of the trace,

$$\begin{aligned} \text{Tr}(\alpha \xi^{j+k}) &= \text{Tr}(\alpha \xi^{j+k-n} \xi^n) \\ &= \text{Tr} \left(\alpha \xi^{j+k-n} \sum_{\ell < n} h_\ell \xi^\ell \right) \\ &= \sum_{\ell < n} h_\ell \text{Tr} \left(\alpha \xi^{j+k-n+\ell} \right). \end{aligned}$$

That is, the $(j+k)$ 'th reverse diagonal depends linearly on the previous n , for $j+k \geq n$. ■

Lemma 13. *Every lf-Kerdock matrix is trace-Kerdock.*

Proof. There are 2^n lf-Kerdock matrices since the top row of n bits enumerates \mathbb{Z}_2^n . There are 2^n trace-Kerdock matrices K_α since the top-left entry in $D_\alpha = (V^T)^{-1} K_\alpha V^{-1}$ enumerates $\mathbb{F}(2^n)$. So there are equal numbers of lf-Kerdock and trace-Kerdock matrices. Above we showed that every trace-Kerdock is a lf-Kerdock. Our statement follows. ■

Thus we have

Theorem 14. *The set of lf-Kerdock matrices is a maximal lf-Kerdock set.*

Henceforth, we refer to lf-Kerdock and trace-Kerdock matrices as “Kerdock matrices.” As above, a Kerdock code is defined from a Kerdock set \mathcal{K} as $\{\varphi_{P,\ell} : P \in \mathcal{K}, \ell \in \mathbb{Z}_2^n\}$.

3.2. Example Kerdock matrix construction. Let $n = 3$. The polynomial $h(t) = 1 + t^2 + t^3 = h_0 + h_2t^2 + t^3$ is a primitive polynomial over \mathbb{Z}_2 of degree 3. A 3×3 Kerdock matrix

$$P = \begin{pmatrix} a_0 & a_1 & a_2 \\ a_1 & a_2 & a_3 \\ a_2 & a_3 & a_4 \end{pmatrix}$$

has five reverse diagonal parameters, a_0, \dots, a_4 . We construct $P \in \mathcal{K}$ by choosing the top row $(a_0 \ a_1 \ a_2)$ arbitrarily, *e.g.*, $(a_0 \ a_1 \ a_2) = (1 \ 1 \ 1)$. The two remaining reverse diagonals a_3 and a_4 are given by

$$a_3 = a_0 + a_2 = 1 + 1 = 0 \quad \text{and} \quad a_4 = a_1 + a_3 = 1 + 0 = 1.$$

This results in the matrix

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

3.3. Properties of Kerdock codes. We now give a lemma that will be useful in Section 5 as well as in its own right.

Lemma 15. *Fix a primitive polynomial h for defining a finite field and for the Kerdock properties. Let P be a symmetric matrix. The following are equivalent:*

- P is Kerdock;
- For all r and s , we have $[r]^T P[s] = [\sqrt{rs}]^T P[\sqrt{rs}] \pmod{2}$;
- For all x, y and z we have $[x]^T P[yz] = [xy]^T P[z] \pmod{2}$.

Proof. First we show that the two algebraic statements are equivalent. Suppose $[x]^T P[yz] = [xy]^T P[z]$ holds for all x, y , and z . Then, given non-zero r and s , put $x = r, y = \sqrt{s/r}$, and $z = \sqrt{rs}$; it follows that $[r]^T P[s] = [\sqrt{rs}]^T P[\sqrt{rs}]$. Conversely, if $[r]^T P[s] = [\sqrt{rs}]^T P[\sqrt{rs}]$ for all r and s , then, given x, y, z , we have $[x]^T P[yz] = [\sqrt{xyz}]^T P[\sqrt{xyz}] = [xy]^T P[z]$, first putting $r = x$ and $s = yz$ and then putting $r = xy$ and $s = z$.

Now, suppose P is Kerdock and fix x, y , and z . By linearity, it suffices to consider $x = \xi^j$ and $z = \xi^k$, for $0 \leq j, k < n$. Because ξ is a multiplicative generator, it suffices to consider $y = \xi$. If $j < n - 1$ and $k < n - 1$, then $[x]^T P[yz] = [xy]^T P[z]$ follows from Hankelness. If $j < n - 1$ and $k = n - 1$, then $[xy]^T P[z] = [\xi^{j+1}]^T P[\xi^{n-1}]$. By the linear feedback Kerdock property, this equals $\sum_{\ell < n} h_\ell [\xi^j]^T P[\xi^\ell]$. By linearity, this is $[\xi^j]^T P[\sum_{\ell < n} h_\ell \xi^\ell]$. By definition of h , this is $[\xi^j]^T P[\xi^n] = [x]^T P[yz]$, as desired. A similar analysis holds if $j = n - 1$ and $k < n - 1$. The case $j = k = n - 1$ follows from symmetry of P .

Conversely, suppose $[x]^T P[yz] = [xy]^T P[z]$ for all x, y , and z . Consider the j 'th row of P , for $j > 0$. We want to show that it is gotten by shifting the $j - 1$ 'st row to the left and setting the rightmost entry of the j 'th row to the appropriate linear combination of the items in the $j - 1$ 'st row. Put $x = \xi^j, y = \xi$, and $z = \xi^k$. Then, for $k < n - 1$, Hankelness (and, therefore, the statement) follows immediately. For $k = n - 1$, we have $yz = \xi^n = \sum_{\ell < n} h_\ell \xi^\ell$, and the statement follows by additivity of the trace. ■

3.4. Kerdock and random variables of limited independence. We first give a definition of limited independence for a family of random variables. This is satisfied by the positions in a random Kerdock codeword.

Definition 16. *A \mathbb{Z}_4 -code is 3.5-wise independent if the distribution on any three positions of a random codeword is uniform on $(\mathbb{Z}_4)^3$ and, conditioned on any three positions, for any fourth position X , we have $\Pr(X = 0) = \Pr(X = 2)$ and $\Pr(X = 1) = \Pr(X = 3)$.*

For example, in our construction, the joint distribution is uniformly random conditioned on the sum of the four positions being 0 or 2 mod 4.

The notion of 3.5-wise independence is useful because it can substitute for 4-wise independence in some cases, even when 3-wise independence cannot. Consider the exponentiated version of a 3.5-wise independent family, so each value is ± 1 or $\pm i$. Then any four random variables W, X, Y, Z satisfy that W, X, Y are independent and, conditioned on W, X, Y , the expectation of Z is 0, because $Z = \pm 1$ uniformly conditioned on Z real and $Z = \pm i$ uniformly conditioned on Z imaginary; the probability that Z is real is arbitrary. Thus, in a family of N 3.5-wise independent random variables, the first four moments agree (up to constant factors) with the moments of a truly random family. We have $E[|\sum X_j|^k] = \Theta(N^{k/2})$ for $k = 0, 2, 4$ and $E[(\sum X_j)^k] = 0$ for $k = 1, 3$. For the 3-wise independent family RM(1), we have $E[|\sum X_j|^4] = \Theta(N^3)$, since, for any triple W, X, Y of variables, there is exactly one fourth variable Z such that $WXYZ \equiv 1$ and all other 4-tuples have zero expectation.

The following had been known, but not previously presented in terms of 3.5-wise independence.

Lemma 17. *For odd $n \geq 3$, the Kerdock code is 3.5-wise independent but not 4-wise independent. For even $n \geq 3$, the Kerdock code is 3-wise but not 3.5-wise independent.*

Lemma 18. *For any $n \geq 3$, the \mathbb{Z}_2 code of Gray-mapped Kerdocks is 4-wise independent.*

4. FAST LIST DECODING OF KERDOCK AND HANKEL CODES

In this section we show how to perform quick list-decoding of the Hankel code. That is, we are given chosen-sampling access to a signal s and a parameter $k < N^c$ for some small $c \geq \Omega(1)$; our goal is to find, with high probability, all Hankel codewords $\varphi_{P,\ell}$ such that $|\langle \varphi_{P,\ell}, s \rangle|^2 \geq (1/k) \|s\|^2$, in time $\text{poly}(k \log(N))$.

Our algorithm is a straightforward generalization of the algorithm of [KM91]. We do not give all the details of this algorithm; instead, we refer the reader to [KM91]. Loosely speaking, the algorithm in [KM91] finds ℓ for which φ_ℓ has large dot product with s by maintaining a set of candidates for the first j bits of ℓ . For $j = 1, 2, 3, \dots, n$, the algorithm extends each candidate from $j - 1$ to j bits in all (two) ways, then tests each new candidate. The tests insure that the number of candidates remains bounded, so the algorithm remains efficient.

Our algorithm will attempt to find first the P matrix of each vector $\varphi_{P,\ell}$ with $|\langle \varphi_{P,\ell}, s \rangle|^2 \geq (1/k) \|s\|^2$. We will call such P and such $\varphi_{P,\ell}$ *heavy* for s . Then the algorithm will find the ℓ part by demodulating out the contribution of P , and using the algorithm in [KM91] to look for heavy RM(1) vectors for $s\varphi_{P,0}^*$, where $\varphi^* = 1/(N\varphi)$ is the componentwise complex conjugate of φ . This strategy relies on the fact that, up to normalization, $\langle \varphi_{P,\ell}, s \rangle = \langle \varphi_{0,\ell}, s\varphi_{P,0}^* \rangle$, so $\varphi_{0,\ell}$ is heavy for $s\varphi_{P,0}^*$ when $\varphi_{P,\ell}$ is heavy for s .

To find P , our algorithm follows the overall structure of [KM91]. For $j \leq n$, we will maintain a set of candidates for the upper-left j -by- j submatrix of P . The candidates will all be Hankel. For each candidate, we will consider extending it to a $(j + 1)$ -by- $(j + 1)$ Hankel matrix, in one of 4 possible ways. We then test each extended candidate in such a way that, with high probability, all true candidates are kept (no false negatives) but the total number of candidates kept is small

enough that our algorithm is efficient. (We describe the retention criterion and test in more detail below.) Much of the [KM91] algorithm works unchanged in our context; we give few comments on those aspects and instead focus on the changes necessary for the Hankel setting and the reasons our algorithm works for Hankel but not for RM(2). In particular, the retention criterion and test we use are similar to that in [KM91] and the guarantee of no false negatives is similar; the main technical work is showing that there are few (true or false) positives in the new context. That is, the analysis is as follows:

- (1) Our algorithm is correct (finds all true candidates), by an analysis similar to [KM91].
- (2) Our algorithm is efficient:
 - As in [KM91], the efficiency of our algorithm reduces to a non-algorithmic and non-probabilistic fact about the number of codewords with large dot product to the signal and the number of extensions of a $(j-1)$ -by- $(j-1)$ candidate to a j -by- j candidate.
 - For the Hankel code in particular, we bound the number of codewords with large dot products and the number of extensions of a single candidate. This is the only part of the proof where we will be formal since this is where our algorithm departs from previous work.

Let us write $P \succeq \tilde{P}$ if \tilde{P} is a square submatrix of P , consisting of the upper left j -by- j corner of P for some j . As in [KM91], we have an ideal testing criterion for submatrices.

Criterion 19. *A testing procedure keeps candidate \tilde{P} iff there exists some n -by- n matrix $P \succeq \tilde{P}$ and some $\ell \in \mathbb{Z}_2^n$ with $|\langle \varphi_{P,\ell}, s \rangle|^2 \geq (1/k) \|s\|^2$. That is, the procedure keeps \tilde{P} iff there exists some unit-norm complex number c , some $P \succeq \tilde{P}$, and some ℓ with $\Re(c \langle \varphi_{P,\ell}, s \rangle) \geq (1/\sqrt{k}) \|s\|$.*

We will gradually rewrite and weaken this criterion in a sequence of variations given below. By “weaken,” we mean that a “weaker” criterion will keep more matrices than a “stronger” criterion. First, for each j , for each string y'' of length $n-j$, and for indeterminate $y' \in \mathbb{Z}_2^j$, define the restriction $(R_{y''} s)$ by $(R_{y''} s)(y') = s(y' y'')$. (Note that, if $\|\varphi\| = 1$, then $R_{y''} \varphi$ is *not* a unit vector. We have $\|R_{y''} \varphi\|^2 = 2^{j-n}$.)

Because $|\varphi_{P,\ell}|$ is constant, if $\langle R_{y''} \varphi_{P,\ell}, R_{y''} s \rangle$ is large, then there must be many (small) contributions. Formally:

Lemma 20. *Suppose $|\varphi| \equiv 2^{-n/2}$, $\|s\| = 1$, and $|\langle \varphi, s \rangle| \geq \sqrt{1/k}$. Then, for each j , there are at least $2^{n-j}/(4k)$ of $y'' \in \mathbb{Z}_2^{n-j}$ such that $|\langle R_{y''} \varphi, R_{y''} s \rangle| \geq (1/\sqrt{4k}) 2^{j-n}$.*

Proof. Suppose not. Let ψ be φ restricted to the $y = y' y''$ with $|\langle R_{y''} \varphi, R_{y''} s \rangle| \geq (1/\sqrt{4k}) 2^{j-n}$, so the support of ψ has size less than $(2^n/(4k))$, and so $\|\psi\|^2 < 1/(4k)$. Then the at-most- 2^{n-j} possible (y'') ’s with $|\langle R_{y''} \varphi, R_{y''} s \rangle| < (1/\sqrt{4k}) 2^{j-n}$ contribute a total of at most $1/\sqrt{4k}$ toward $|\langle \varphi, s \rangle|$, i.e., $|\langle \varphi - \psi, s \rangle| \leq 1/\sqrt{4k}$. It follows that

$$\begin{aligned}
|\langle \varphi, s \rangle| &\leq |\langle \psi, s \rangle| + |\langle \varphi - \psi, s \rangle| \\
&\leq \|\psi\| \|s\| + 1/\sqrt{4k} \\
&< 1/\sqrt{4k} + 1/\sqrt{4k} \\
&= 1/\sqrt{k},
\end{aligned}$$

a contradiction. ■

Thus we can weaken Criterion 19 to:

Criterion 21. *A testing procedure keeps candidate \tilde{P} iff there exists some n -by- n matrix $P \succeq \tilde{P}$, a unit-magnitude complex number c , and some $\ell \in \mathbb{Z}_2^n$ such that, for at least $2^{n-j}/(4k)$ of $y'' \in \mathbb{Z}_2^{n-j}$ we have $\Re(c \langle R_{y''} \varphi_{P,\ell}, R_{y''} s \rangle) \geq (1/(\sqrt{4k})) 2^{j-n} \|s\|$.*

Next, weaken Criterion 21 to

Criterion 22. A testing procedure keeps candidate \tilde{P} iff there exists some n -by- n matrix $P \succeq \tilde{P}$ and, for at least $2^{n-j}/(4k)$ of $y'' \in \mathbb{Z}_2^{n-j}$ there exists some unit-norm $c_{y''}$ and some $\ell_{y''} \in \mathbb{Z}_2^{n-j}$ with $\Re \left(c_{y''} \left\langle R_{y''} \varphi_{P, \ell_{y''}}, R_{y''} s \right\rangle \right) \geq (1/(\sqrt{4k})) 2^{j-n} \|s\|$.

Next, we will show that we need not search over all possible $P \succeq \tilde{P}$; a single fixed extension \tilde{P}' will suffice. (For example, \tilde{P}' might extend \tilde{P} with zeros; \tilde{P}' need not even be Hankel.) We will use the notation \tilde{P}' in the sequel.

Lemma 23. Fix parameter n , $j < n$, j -by- j matrix \tilde{P} , extension $\tilde{P}' \succeq \tilde{P}$, and $y'' \in \mathbb{Z}_2^{n-j}$. Then

$$\{R_{y''} \varphi_{P, \ell} : P \succeq \tilde{P}, \ell \in \mathbb{Z}_2^n\} = \{R_{y''} \varphi_{\tilde{P}', \ell} : \ell \in \mathbb{Z}_2^n\}.$$

Proof. Write an extension P to \tilde{P} as

$$P = \begin{pmatrix} \tilde{P} & P_1^T \\ P_1 & P_2 \end{pmatrix}$$

and write $\ell^T = (\ell_1^T | \ell_2^T)$, where $\ell_1 \in \mathbb{Z}_2^j$. Then, at $y = y' y''$, we have

$$\begin{aligned} y^T P y + 2\ell^T y &= (y')^T \tilde{P} y' + 2(y'')^T P_1 y' + (y'')^T P_2 y'' + 2\ell_1^T y' + 2\ell_2^T y'' \\ &= (y')^T \tilde{P} y' + 2((y'')^T P_1 + \ell_1^T) y' + ((y'')^T P_2 y'' + 2\ell_2^T y''). \end{aligned}$$

If we fix y'' but let ℓ vary, the expression $2((y'')^T P_1 + \ell_1^T)$ varies over all of $2\mathbb{Z}_4^j$, whether or not we let P_1 vary. Similarly, if we fix y'' but let the coefficient c vary, the expression $c i^{(y'')^T P_2 y'' + 2\ell_2^T y''}$ varies over unit-norm complex numbers, whether or not we let P_2 (and ℓ_2) vary. ■

It follows that we can rewrite Criterion 21 as

Criterion 24. A testing procedure keeps candidate \tilde{P} iff for at least $2^{n-j}/(4k)$ of $y'' \in \mathbb{Z}_2^{n-j}$ there exists some unit-norm $c_{y''}$ and some $\ell_{y''} \in \mathbb{Z}_2^{n-j}$ with

$$\Re \left(c_{y''} \left\langle R_{y''} \varphi_{\tilde{P}', \ell_{y''}}, R_{y''} s \right\rangle \right) \geq (1/(\sqrt{4k})) 2^{j-n} \|s\|.$$

Finally, we will not be able to compute the test exactly, but we will approximate with samples. To that end, we need to have two thresholds, with a gap. Formally, we want the following criterion, in which both the first and third cases represent a weakening, compared with Criterion 24:

Criterion 25. A testing procedure of a j -by- j Hankel matrix \tilde{P} and signal s with parameters c_1 and c_2 (determined below) behaves as follows.

- If for at least $2^{n-j}/(4k)$ of $y'' \in \mathbb{Z}_2^{n-j}$ there exists some unit-norm $c_{y''}$ and some $\ell_{y''} \in \mathbb{Z}_2^{n-j}$ with $\Re \left(c_{y''} \left\langle R_{y''} \varphi_{\tilde{P}', \ell_{y''}}, R_{y''} s \right\rangle \right) \geq (1/(\sqrt{4k})) 2^{j-n} \|s\|$, the procedure keeps \tilde{P} with high probability.
- If only for less than $c_1 2^{n-j}/(4k)$ of $y'' \in \mathbb{Z}_2^{n-j}$ does there exist some unit-norm $c_{y''}$ and some $\ell_{y''} \in \mathbb{Z}_2^{n-j}$ with $\Re \left(c_{y''} \left\langle R_{y''} \varphi_{\tilde{P}', \ell_{y''}}, R_{y''} s \right\rangle \right) \geq (1/(c_2 \sqrt{4k})) 2^{j-n} \|s\|$, the procedure drops \tilde{P} with high probability.
- (The procedure may behave arbitrarily, otherwise.)

Our algorithm will also need an estimate for $\|s\|$. Here we simply assume that $\|s\|$ is known, say, up to the factor 2. Alternatively, one might assume the *dynamic range* of the problem is bounded, i.e., that $1/M \leq \|s\| \leq M$ for some known M . The algorithm could then try all $O(\log(M))$ possible 2^j in the range $1/M$ to M ; one of them is a factor-2 approximation to $\|s\|$. This leads to an extra

factor of $\log(M)$ in some costs. One can also get an appropriate approximation to $\|s\|$ from samples to s without an assumption about the dynamic range. We omit details; see [GGI⁺02].

We use the following straightforward efficient sampling algorithm to implement Criterion 25, for which there exist suitable c_1 and c_2 :

Algorithm 26. Assuming $\|s\|$ is known to within a constant factor (that is absorbed into c_2):

- For each $y'' \in \mathbb{Z}_2^{n-j}$ such that $\|R_{y''}s\|^2 \leq (40k/c_1)2^{j-n}\|s\|^2$, we can use the [KM91] algorithm to determine whether $\ell_{y''}$ and $c_{y''}$ for Criterion 25 exist.
- To determine whether at least $2^{n-j}/(4k)$ or at most $c_1 2^{n-j}/(4k)$ of the $y'' \in \mathbb{Z}_2^{n-j}$ satisfy our condition, sample approximately k/c_1 of the y'' 's; repeat to drive down failure probability. Note that there are at most $(c_1/10)2^{n-j}/(4k)$ possible (y'') 's where $\|R_{y''}s\|^2 > (40k/c_1)2^{j-n}\|s\|^2$. The algorithm can behave arbitrarily on these y'' and still distinguish “at most $(c_1)2^{n-j}/(4k)$ ” from “at least $2^{n-j}/(4k)$.”

■

In summary, the following is a direct generalization of previous work on RM(1) (*e.g.*, [KM91]) concerning false negatives, for which there is nothing special about RM(1) or Hankel:

Proposition 27. Fix parameter n and signal s of length $N = 2^n$.

- For any k and any $j \leq n$, any procedure satisfying Criterion 25 keeps, with high probability, all j -by- j Hankel matrices \tilde{P} for which there is some $P \succeq \tilde{P}$ and some $\ell \in \mathbb{Z}_2^n$ with $|\langle s, \varphi_\lambda \rangle|^2 \geq (1/k)\|s\|_\sim^2$.
- Algorithm 26 satisfies Criterion 25.
- Algorithm 26 runs in time $\text{poly}(k \log(N))$.

Thus we have shown that *each* call to Algorithm 26, to test a *single* candidate, is efficient. We will call Algorithm 26 on many candidates as follows.

Algorithm 28. Start with the exhaustive candidate set C_1 for 1-by-1 matrices \tilde{P} . For j increasing from 1 to $n - 1$, extend all candidates in C_j to $(j + 1)$ -by- $(j + 1)$ Hankel matrices in all possible ways. Call Algorithm 26 to test each candidate extension.

■

It remains to show that the number of candidates \tilde{P} under consideration remains under control. Let $f(j)$ denote the number of j -by- j candidates considered. Each candidate will be extended to a $(j + 1)$ -by- $(j + 1)$ Hankel matrix in all possible ways, getting $g(j + 1)$ possible $(j + 1)$ -by- $(j + 1)$ candidates. Then Criterion 25 will be applied to each candidate, reducing the number of candidates from $g(j + 1)$ to $f(j + 1)$. We need to bound both $f(j)$ and $g(j)$. We first bound $g(j + 1)$ by $4f(j)$:

Lemma 29. Algorithm 28 constructs only four extensions to any candidate.

Proof. Note that a j -by- j candidate \tilde{P} extends to $(j + 1)$ -by- $(j + 1)$ in only four ways, since there are only two new possible bits, a and b :

$$\left(\begin{array}{c|c} \tilde{P} & \\ \hline a & b \end{array} \right).$$

■

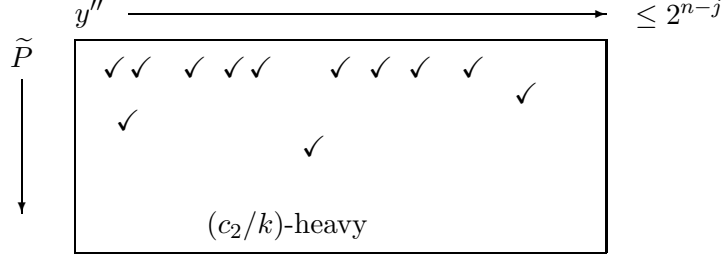


FIGURE 1. Bounding the number of \tilde{P} 's that are heavy on many (y'') 's. Label columns by (y'') 's for which $\|R_{y''}s\|^2 \leq (k/c_3)(2^{j-n})\|s\|^2$ and label rows by \tilde{P} 's; put a checkmark at (y'', \tilde{P}) if \tilde{P} is (c_2/k) -heavy for y'' . We will show below that there are few checkmarks in any column; it follows that there are few rows with checkmarks in many columns.

Here we crucially use the fact that the candidates are Hankel. If we were to consider arbitrary j -by- j RM(2) matrices, the number of extensions would be 2^{j+1} , which is prohibitive.

Thus it suffices to bound $f(j)$ by a polynomial in k , uniformly for all j . So we need to bound the number of \tilde{P} 's that are (c_2/k) -heavy on more than $c_1(k)2^{n-j}/k$ of the $y'' \in \mathbb{Z}_2^{n-j}$, where we call a candidate \tilde{P} *h-heavy on y''* if \tilde{P} extends to some P with some $\ell_{y''}$ satisfying $\left| \langle R_{y''} \varphi_{P, \ell_{y''}}, R_{y''} s \rangle \right| \geq \sqrt{h} 2^{j-n} \|s\|$. The other candidates are dropped by our criterion.

As in previous work, it suffices to bound the number of candidates \tilde{P} for *each* y'' and then do an averaging argument. There are at most $(c_3/k)2^{n-j}$ possible (y'') 's for which $\|R_{y''}s\|^2 \geq (k/c_3)2^{j-n}\|s\|^2$ and, for constant c_3 related to c_1 , these (y'') 's can be ignored in determining whether \tilde{P} satisfies the condition of Criterion 25 on at least $(1/(4k))2^{n-j}$ or at most $(c_1/(4k))2^{n-j}$ of the (y'') 's. So, henceforth, consider only y'' for which $\|R_{y''}s\|^2 < (k/c_3)2^{j-n}\|s\|^2$. Below, for all such y'' , we will bound, by $B_k \leq \text{poly}(k)$, the number of \tilde{P} that are (c_2/k) -heavy on y'' . Summing over at most 2^{n-j} possible (y'') 's, there are at most $B_k 2^{n-j}$ pairs (\tilde{P}, y'') where \tilde{P} is (c_2/k) -heavy for y'' . Thus there can be at most $B_k \cdot (k/c_1) \leq \text{poly}(k)$ possible \tilde{P} 's that are (c_2/k) -heavy on at least $(c_1/k)2^{n-j}$ of the (y'') 's; *i.e.*, at any stage j , there are at most $f(j) \leq \text{poly}(k)$ possible candidates considered by our algorithm. See Figure 1.

Thus we have, from previous work and without specific consideration of the Hankel code,

Proposition 30. *Fix signal s of length $N = 2^n$, fix parameter k , and fix $j \leq n$. Suppose, for each $y'' \in \mathbb{Z}_2^{n-j}$ with $\|R_{y''}s\|^2 \leq (k/c_3)\|R_{y''}s\|^2$, there are at most $\text{poly}(k)$ possible j -by- j Hankel matrices \tilde{P} that are (c_2/k) -heavy for y'' . Then there are at most $\text{poly}(k)$ possible \tilde{P} that are kept by our algorithm.*

Finally, we now proceed to Hankel-specific analysis. To simplify notation, and without loss of generality, we drop all previous constants. It suffices to show that there are at most $\text{poly}(k)$ Hankel matrices P such that there exists an ℓ with $|\langle s, \varphi_{P, \ell} \rangle|^2 \geq (1/k)\|s\|^2$.

In an orthonormal basis, by the Parseval Equality, there can be at most k vectors φ with $|\langle s, \varphi \rangle|^2 \geq (1/k)\|s\|^2$. Similarly, in a μ -incoherent dictionary, *i.e.*, a set of vectors with all dot product magnitudes bounded above by μ , if μk is at most some constant $c_4 \approx 1/6$, then there are at most $O(k)$ such λ 's [TGMS03, GMS03]. Hankel, however, is not a μ -incoherent set for small μ , because there are pairs P and P' of Hankel matrices that differ by a low-rank matrix, whence the corresponding vectors $\varphi_{P, 0}$ and $\varphi_{P', 0}$ have large dot product. Nevertheless, we show that, for each

P , the number of P' such that $P + P'$ has low rank is small. We then show that the set of Hankel codewords works like an orthonormal basis or an incoherent set, in the sense that there may be at most $\text{poly}(k)$ Hankel λ 's with $|\langle s, \varphi_\lambda \rangle|^2 \geq (1/k) \|s\|^2$.

We now proceed formally. This proceeds in a sequence of lemmas along with Dickson's Theorem (Theorem 6), all of which have proofs that are elementary or found in existing work.

Lemma 31. *There is a constant c_4 such that, for any incoherence parameter μ , $0 \leq \mu \leq 1$ any k , and any signal s , if $\mu k \leq c_4$, there are at most $O(k)$ vectors in any set A such that both of the following hold:*

- For all $\varphi \neq \varphi' \in A$, we have $|\langle \varphi, \varphi' \rangle| \leq \mu$.
- For all $\varphi \in A$, we have $|\langle s, \varphi \rangle|^2 \geq (1/k) \|s\|^2$.

Proof. This essentially follows from [TGMS03, GMS03]; we include a sketch of the proof with possibly different constants. Suppose, toward a contradiction, there are $\ell > 4k$ vectors in A ; wlog, $\ell = 4k + 1$, since we can discard the remaining vectors. We may assume that $s = \sum_j a_j \varphi_j$ lies in the span of $A = \{\varphi_j\}$. The idea is to show that $\|s\|^2 \approx \sum_j |a_j|^2$ and $|\langle s, \varphi_j \rangle|^2 \approx |a_j|^2$, so that an approximate Parseval equality holds. First,

$$\begin{aligned} \|s\|^2 &= \left\langle \sum_j a_j \varphi, \sum_{j'} a_{j'} \varphi \right\rangle \\ &\geq \sum_j |a_j|^2 - \mu \left| \sum_{j \neq j'} a_j \overline{a_{j'}} \right| \\ &\geq \sum_j |a_j|^2 - \mu \left| \sum_j a_j \right|^2 \\ &\geq \sum_j |a_j|^2 - \mu(4k+1) \sum_j |a_j|^2, \end{aligned}$$

by Cauchy-Schwarz, so that, for some c ,

$$\sum_j |a_j|^2 \leq (1 + c\mu k) \|s\|^2. \quad (2)$$

On the other hand, for each j ,

$$\begin{aligned} |\langle s, \varphi_j \rangle| &= \left| a_j + \sum_{j' \neq j} a_{j'} \langle \varphi_{j'}, \varphi_j \rangle \right| \\ &\leq |a_j| + \mu \sum_{j' \neq j} |a_{j'}| \\ &\leq |a_j| + \mu \sqrt{4k \sum_{j' \neq j} |a_{j'}|^2} \\ &= |a_j| + O(\mu k)(1/\sqrt{k}) \|s\|, \end{aligned}$$

so that, for some c' we have $|a_j| \geq |\langle s, \varphi_j \rangle| - O(\mu k)(1/\sqrt{k}) \|s\|$, and so

$$|a_j|^2 \geq (1 - c'\mu k)^2 (1/k) \|s\|^2.$$

Summing over all $\ell = 4k + 1$ terms, we get

$$\sum_j |a_j|^2 \geq (1 - c'\mu k)^2 (\ell/k) \|s\|^2,$$

so that, with (2), we get $(1 - c'\mu k)^2 (\ell/k) \leq (1 + c\mu k)$, or $\ell \leq k(1 + c\mu k)(1 - c'\mu k)^{-2}$. Thus, if $k\mu$ is a sufficiently small constant, we get $\ell \leq 4k$, a contradiction. ■

Lemma 32. *For any constant c_4 , there are just $L_k \leq \text{poly}(k)$ Hankel matrices of rank at most $2\log(k/c_4)$. Equivalently, for each Hankel P , there are at most L_k Hankels P' with $\text{rank}(P + P') \leq 2\log(k/c_4)$.*

Proof. Suppose Hankel matrix P has rank r . We claim that $O(r)$ binary parameters determine the top half of the matrix (above the main reverse diagonal). Another $O(r)$ parameters determine the bottom half, whence the number of such matrices is $2^{O(r)}$. The result follows.

Write the $(r+1)$ 'st column as a linear combination C of the first r columns. We claim that C and the first r entries p_0, p_1, \dots, p_{r-1} in the top row determine the top half of the matrix. Determine p_r from p_0, p_1, \dots, p_{r-1} and C applied to the top row (row 0). Then, having determined p_r , determine p_{r+1} from C applied to the first r entries in row 1, *i.e.*, p_1, p_2, \dots, p_r . Proceed to determine p_{r+2} from C applied to the first r entries in row 2, *i.e.*, p_2, p_3, \dots, p_{r+1} . The general statement follows by induction.

For example, suppose Hankel P has rank three, the first three reverse diagonal parameters are a, b, c , and column 3 is the linear combination C of columns 0, 1, 2. Then, in

$$P = \left(\begin{array}{ccc|c|c} a & b & c & d & e \\ b & c & d & e & f \\ c & d & e & f & g \\ d & e & f & g & h \\ \vdots & & & & \end{array} \right),$$

we get d in row 0, column 3 from a, b, c by applying C in row 0. Now knowing d in addition to a, b, c , we get e in row 1, column 3 by applying C to b, c, d in row 1. We get f in row 2, column 3 by applying C to c, d, e , etc. ■

In intermediate stages of our algorithm, we need to bound only the number of Hankel *matrices* P that are considered. In the output, however, we need to bound the total number of Hankel *codewords* output, *i.e.*, the number of pairs (P, ℓ) . We give the latter stronger statement in this summary theorem.

Theorem 33. *For any signal s , there are at most $\text{poly}(k)$ Hankel codewords $\varphi_{P,\ell}$ with*

$$|\langle s, \varphi_{P,\ell} \rangle|^2 \geq (1/k) \|s\|^2.$$

Proof. Suppose there are at least q Hankel codewords $\varphi_{P,\ell}$ with $|\langle s, \varphi_{P,\ell} \rangle|^2 \geq (1/k) \|s\|^2$. For fixed P , the set $\{\varphi_{P,\ell} : \ell\}$ is an orthonormal basis, so there are at most k possible ℓ 's for each P with $|\langle s, \varphi_{P,\ell} \rangle|^2 \geq (1/k) \|s\|^2$. Thus there are at least q/k matrices P with at least one ℓ satisfying $|\langle s, \varphi_{P,\ell} \rangle|^2 \geq (1/k) \|s\|^2$. By Lemma 32, there is a set Q of size $|Q| \geq q/(kL_k)$ matrices P having an ℓ satisfying $|\langle s, \varphi_{P,\ell} \rangle|^2 \geq (1/k) \|s\|^2$ and with $\text{rank}(P + P') \geq 2\log(k/c_4)$ for all $P \neq P' \in Q$. By Theorem 6, for any $P \neq P' \in Q$ and their corresponding ℓ and ℓ' , we have $|\langle \varphi_{P,\ell}, \varphi_{P',\ell'} \rangle| \leq (c_4/k)$. By Lemma 31, $|Q| \leq O(k)$. It follows that $q \leq \text{poly}(k)$. ■

In summary, we have our main theorem.

Theorem 34. *Let $\{\varphi_\lambda\}$ denote the Hankel code. There is an algorithm that, given parameter k and chosen-sampling access to a signal $s \in \mathbb{C}^N$, finds, in time $\text{poly}(k \log(N))$, a list containing all λ with $|\langle s, \varphi_\lambda \rangle|^2 \geq (1/k) \|s\|^2$.*

5. CONCLUSION

5.1. Corollaries. A list-decoding algorithm for the Hankel code immediately gives a list-decoding algorithm for the Kerdock subcode. Since the Kerdock code is $(1/\sqrt{N})$ -incoherent, we immediately get a sparse recovery algorithm for Kerdock, using [TGMS03, GMS03]. That is:

Corollary 35. *Let $\{\varphi_\lambda\}$ denote a Kerdock code that is a subset of a Hankel code. There is an algorithm that, given parameter k and chosen-sampling access to a signal $s \in \mathbb{C}^N$, finds, in time $\text{poly}(k \log(N))$, a list containing all λ with $|\langle s, \varphi_\lambda \rangle|^2 \geq (1/k) \|s\|^2$.*

Corollary 36. *Let $\{\varphi_\lambda\}$ denote a Kerdock code that is a subset of a Hankel code. There is an algorithm that, given parameters $k < 1/(6\sqrt{N})$ and $\epsilon > 0$ and chosen-sampling access to a signal $s \in \mathbb{C}^N$, finds, in time $\text{poly}(k \log(N)/\epsilon)$, a set Λ of size k and coefficients c_λ (i.e., a k -term approximation $\tilde{s} = \sum_{\lambda \in \Lambda} c_\lambda \varphi_\lambda$) with $\|\tilde{s} - s\|^2 \leq (1 + \epsilon + k^2/\sqrt{N}) \|s_k - s\|^2$, where s_k is the best k -term Kerdock approximation to s .*

5.2. Improvements. The cost of our Hankel recovery algorithm is polynomial in k , but high. In Lemma 32, we show only that there are at most 2^{4r} Hankel matrices of rank r , whence, for each Hankel P , there are at most $2^{4r} = k^8$ Hankel matrices $P' \neq P$ with $|\langle \varphi_{P,\ell}, \varphi_{P',\ell'} \rangle| > (1/k) = 2^{-r/2}$. This means we bound the time cost of our algorithm at k^c for c an integer somewhat larger than 8. We make a few comments:

- It is easy to see that there are at least $\Omega(k^4)$ Hankel matrices of rank $1/k$. If we really want to list-decode Hankel rather than Kerdock, the size of the output can really be at least approximately k^5 . Our runtime of k^c will be approximately quadratic in the size of the output, which may be acceptable in some contexts.³
- A tighter analysis of the way the top and bottom halves of the matrix fit together may bound the number of rank- $(1/k)$ Hankels more tightly than k^8 .
- We have begun to investigate an alternative algorithm that exploits the fact that the restriction of a Kerdock codeword to a subfield is a smaller instance of a Kerdock codeword. This algorithm is much faster as a list-decoding algorithm for Kerdock only, since it doesn't keep so many candidates. But the paradigm of bit-by-bit extensions in the algorithm of [KM91] and Algorithm 28 does not work for subfields.

Faster algorithms to list-decode Kerdock codes will be the subject of future work.

Other future work will include extensions to the Delsarte-Goethals hierarchy of codes between $\text{RM}(1)$ and $\text{RM}(2)$. As one ascends the hierarchy, the size of the code increases as the the maximum dot product increases.

ACKNOWLEDGMENT

We thank Joel Lepak, Muthu Muthukrishnan and Alex Samorodnitsky for helpful discussions.

³The $\approx k^5$ output Hankel codewords come in (possibly overlapping) clusters of approximately k^4 vectors each, so there are at most approximately k clusters. One might hope to produce a compressed representation of the output in less time than it takes to write out the output uncompressed. Note, however, that the boundaries of the clusters are generally not smooth, so it will not suffice to output the cluster centers.

REFERENCES

- [AGS03] A. Akavia, S. Goldwasser, and S. Safra. Proving hard-core predicates using list decoding. In *in Proc. of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.
- [AKK⁺03] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing low-degree polynomials over $gf(2)$. In *Proc. of APPROX+RANDOM*, 2003.
- [AS97] Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. In *Proc. 29th ACM symposium on Theory of computing*, pages 485–495, 1997.
- [CKKS97] A.R. Calderbank, P.J. Cameron, W.M. Kantor, and J.J. Seidel. \mathbb{Z}_4 -kerdock codes, orthogonal spreads, and extremal euclidean line-sets. *Proc. London Math. Soc. (75)*, pages 436–480, 1997.
- [CGL⁺05] A. R. Calderbank, A. Gilbert, K. Levchenko, S. Muthukrishnan, and M. Strauss. Improved range-sumnable random variable construction algorithms. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 840–849, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.
- [GGI⁺02] Anna C. Gilbert, Sudipto Guha, Piotr Indyk, S. Muthukrishnan, and Martin Strauss. Near-optimal sparse Fourier representations via sampling. In *Proc. 34th ACM Symposium on Theory of Computing*, pages 152–161, 2002.
- [GL89] O. Goldreich and L. Levin. A hard-core predicate for all one-way functions. In *Proc. 21st ACM Symposium on Theory of Computing*, pages 25–32. ACM, 1989.
- [GMS03] A. Gilbert, S. Muthukrishnan, and M. Strauss. Approximation of functions over redundant dictionaries using coherence, 2003.
- [HCM06] S. D. Howard, A. R. Calderbank, and W. Moran. The finite heisenberg-weyl groups in radar and communications. *EURASIP Journal on Applied Signal Processing*, 2006:Article ID 85685, 12 pages, 2006.
- [HKC⁺94] A. R. Hammons, P.V. Kumar, A. R. Calderbank, N.J.A. Sloane, and P. Solé. The \mathbb{Z}_4 -linearity of Kerdock, Preparata, Goethals, and related codes. *IEEE Trans. on Information Theory*, 40(2):301–319, 1994.
- [HSP06] R. W. Heath, T. Strohmer, and A. J. Paulraj. On quasi-orthogonal signatures for CDMA systems. *IEEE Transactions on Information Theory*, 52(3), March 2006.
- [Ind00] P. Indyk. *High-Dimensional Computational Geometry*. PhD thesis, Stanford, 2000.
- [KL05] T. Kaufman and S. Litsyn. Almost orthogonal linear codes are locally. In *Proc. 46th Foundations of Computer Science*, pages 317–326. IEEE, 2005.
- [KM91] Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the Fourier spectrum. pages 455–464, 1991.
- [Man95] Y. Mansour. Randomized interpolation and approximation of sparse polynomials. *SIAM Journal on Computing*, 24(2):357–368, 1995.
- [MS77] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, 1977.
- [Sud] Madhu Sudan. Algorithmic introduction to coding theory. Course Home Page. <http://theory.lcs.mit.edu/~madhu/FT01/>.
- [Sud00] Madhu Sudan. List decoding: Algorithms and applications. In *Proc. of the International Conference IFIP TCS 2000*, 2000.
- [Sud01] Madhu Sudan. Coding theory: Tutorial and survey. In *Proc. of the 42nd Annual Symposium on Foundations of Computer Science*, 2001.
- [TGMS03] J. A. Tropp, A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss. Improved sparse approximation over quasi-incoherent dictionaries. In *Proc. of the 2003 IEEE International Conference on Image Processing*, Barcelona, 2003.

E-mail address: calderbk@math.princeton.edu, {annacg,martinjs}@umich.edu

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, FINE HALL, PRINCETON, NJ 08544

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, 2074 EAST HALL, ANN ARBOR, MI 48109

DEPTS. OF MATHEMATICS AND EECS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109